5.1 A Power-Efficient High-Throughput 32-Thread SPARC Processor

Ana Sonia Leon, Jinuk Luke Shin, Kenway W. Tam, William Bryg, Francis Schumacher, Poonacha Kongetira, David Weisner, Allan Strong

SUN Mircrosystems, Sunnyvale, CA

The 64b Niagara SPARC processor is designed for power-efficient high-throughput commercial server applications where power, cooling, and space are major concerns [1, 2]. The chip-multithreaded (CMT) architecture achieves high throughput while optimizing performance/watt. Concurrent execution of 32 threads is implemented through 8 symmetrical 4-way multithreaded cores, supported by a high-bandwidth low-latency cache/memory system (Fig. 5.1.1). Each core has a simple single-issue 6-stage pipeline where instructions from all 4 threads are interleaved per cycle with zero thread-switch cost, maximizing pipeline utilization. When any thread is blocked by a cache miss or branch penalty, the other threads issue instructions more frequently, effectively hiding the miss latency of the first thread. Cache/memory latency is minimized using sufficient bandwidth and physical proximity: CPU-to-cache crossbar of 134GB/s, 4-banked 12-way pipelined shared 3MB L2 cache of 153.6GB/s, and 4 double-width DDR2 DIMM channels at 400MT/s (Mega-transfers/s) delivering 25.6GB/s. The result is a measured IPC (instructions per cycle) of 5.76 with an actual L2 latency of 20.9 CPU cycles and memory latency of 106ns on Java Business Benchmark (SpecJBB), for a pipeline efficiency of 71% (5.76 out of a maximum of 8). The chip is implemented in a 90nm CMOS process with 9 layers of Cu interconnect. The 378mm^2 die (Fig. 5.1.7) comprises 279M transistors, packaged in a flip-chip ceramic LGA with 1933 pins. Power dissipation is 63W at 1.2V and 1.2GHz (Fig 5.1.2), leading to high performance/watt.

The power-efficient architecture enables a massively parallel design at a less aggressive frequency, with high resource utilization. Power-hungry techniques like memory speculation, out-of-order execution, and predication are not needed to achieve the desired performance. Active power- and temperature-control mechanisms allow threads and cores to be dynamically scheduled or idled, keeping the peak power close to the average power. Clock-gating techniques include coarse-grain to disable selective cores, and fine-grain to disable about 30% of the datapath flops on average. A balanced H-tree scheme is used to distribute the global clock. Library cells are static CMOS with a 1.5 P/N width ratio to lower gate capacitance. Interconnect power is minimized by optimizing wire classes for power-delay product. The CMT use of 8 identical cores results in a more uniform power distribution across the chip with a thermal gradient of only 7°C. At 63W, the measured worst-case junction temperature is 66°C. Compared to a typical Tj of 105°C, reliability improves by 5×, thereby decreasing implementation complexity due to electromigration (EM), gate-oxide integrity (GOI), negative bias temperature instability (NBTI), and channel hot carrier (CHC) along with providing an opportunity for supply-voltage overdrive.

This CMT design approach achieves high throughput without requiring high frequency, minimizing implementation complexity and shortening the design cycle. Extensive use of simple static CMOS circuits improves the robustness. Custom cell design is only needed for the analog circuits, the SRAMs, the initial receive and final drive stages of the IOs. A simple and flexible clocking scheme uses a single PLL to generate 3 ratioed clock domains: (1) CPU, crossbar, L2 cache; (2) memory interface; (3) system interface. The last stage of the design cycle is shortened by a hold-time methodology based on metal-programmable delay buffers, allowing the top level route to freeze while still resolving violations.

The integer register file (IRF) in each core supports 4 threads while achieving high area- and power-efficiency. A custom highly integrated memory cell structure is designed to support the 8 register windows typical of SPARC architecture processors. A high-speed bidirectional differential port enables the transfer of data between an active window storage element and the eight base window cells. The skew-tolerant design of this port allows great variability of timing between 'wlb' and 'save' signals (Fig 5.1.3), ensuring the ability to write despite parasitic read on transfer bitlines (blt). The new memory cell, in combination with register renaming (in/out) logic, eliminates unnecessary data movement and associated power consumption. Since the pipeline design is single-issue, the 4 threads can share the 3 read ports needed for an active thread, achieving high array density and reducing routing congestion.

The row decoder (Fig 5.1.4) is implemented with an internal pipeline to stage back-to-back swap requests. This allows 2 consecutive swaps to overlap in time, maintaining single-cycle throughput. External register-management control logic prevents conflicts between requests while power density is managed by limiting swaps to 16 registers at a time. Integrating FFs and low phase transparent latches ("B" latches) in the final decoder allows the 3-cycle swap operation to be orchestrated using locally staged save and restore control signals. Concurrently, the final decoder frees up the pre-decoder for the next swap request, to generate current window pointers and thread IDs.

Another important feature of the IRF is an integrated single-ended sense amplifier latch with a built-in 2-to-1 column decoder to facilitate high-frequency operation (Fig 5.1.5). The footer device 'MSEL' is used to disable the receiver and clear the latch output, allowing the other selected column data to propagate through the output driver 'ND2' without any delay penalty.

The 12-way set-associative on-chip L2 cache is divided into 4 independent banks that can operate concurrently to read out up to 256B. Each bank includes data and tag SRAMs, directory CAMs, a valid-used-allocate-dirty (VUAD) register file, and control logic. The banks are further divided into 4 sub-banks of 12 panels, each containing one of the data ways. Each sub-bank supplies 16B with 2-cycle throughput, providing a maximum data array read bandwidth of 153.6GB/s. Extensive gating schemes are implemented, enabling 1 to 4 of the 48 panels in the bank to be activated according to the requested data size. A one-hot way select signal is used to activate the individual panels in a sub-bank. To provide clock gating and minimize global clock distribution, special level-2 (cluster) clock headers are shared by 2 panels in a sub-bank. Headers are enabled only when an access is requested to the sub-bank. The clock header circuit, Fig 5.1.6, also implements an inter-locking scheme to prevent back-to-back requests to the same sub-bank. Since the number of L1 lines is lower than the number of L2 lines, using L1 cache tags to manage cache coherency in the large L2 directory CAM arrays reduces both area and power. In addition, indices are arranged for power efficiency, so that each access enables only 1/8th of this reverse-mapped directory CAM.

Acknowledgments:
The authors acknowledge the contributions from the Niagara Development Team at Sun, and TI for fabrication.

References:
[1] P. Kongetira, "A 32-Way Multithreaded SPARC Processor," *16th Hot Chips Symp.*, Aug., 2004.
[2] P. Kongetira, et al, "A 32-Way Multithreaded SPARC Processor," *IEEE Micro*, vol. 25, pp. 21-29, Mar., 2005.

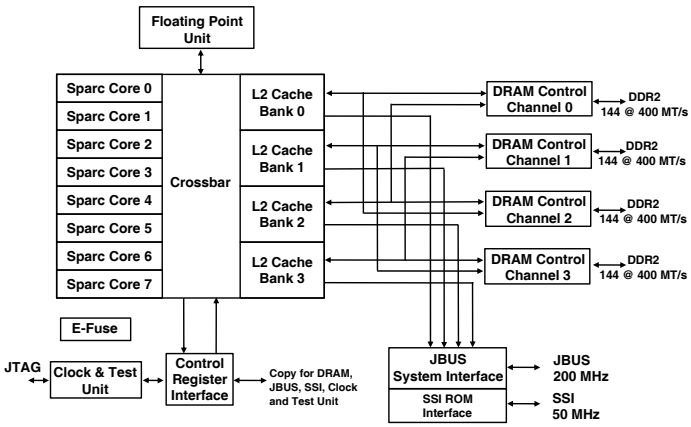


Figure 5.1.1: Processor block diagram.

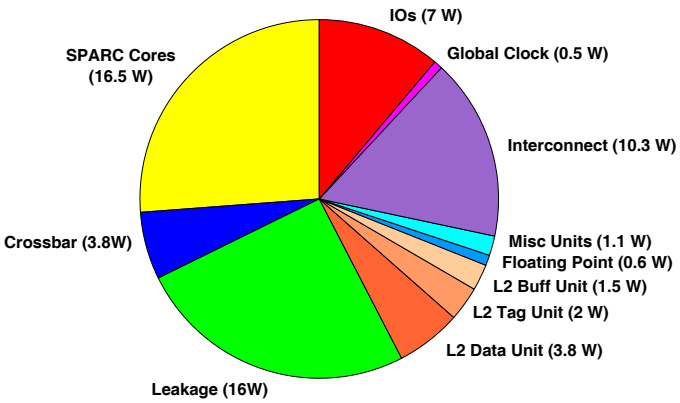


Figure 5.1.2: Chip power consumption: 63W.

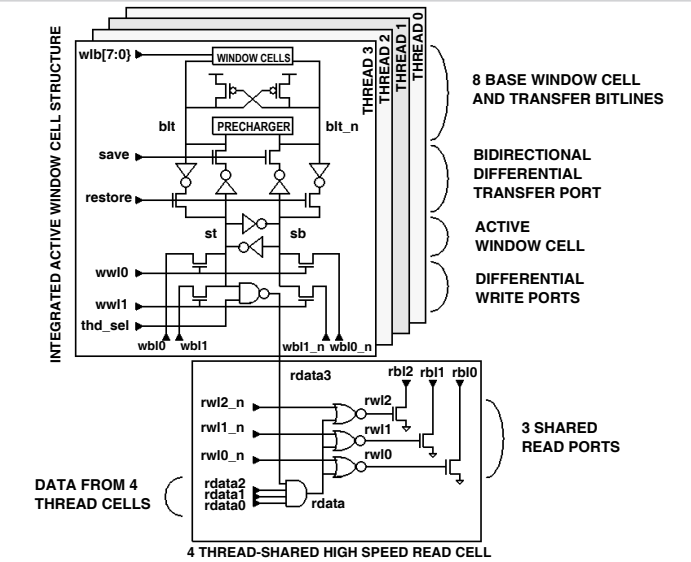


Figure 5.1.3: Integer register file memory cell structure.

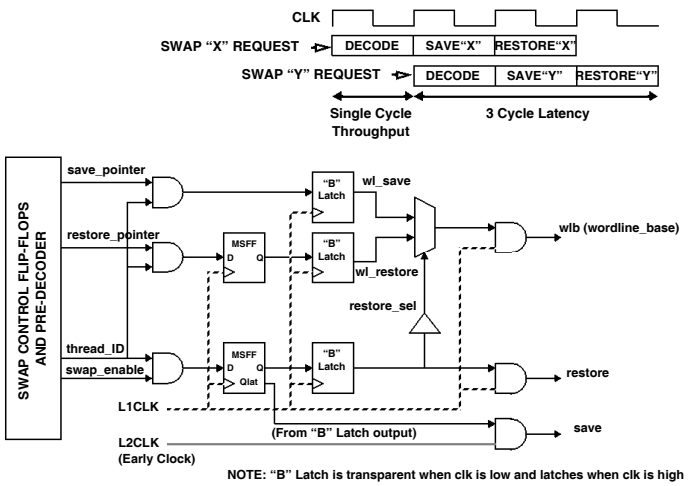


Figure 5.1.4: IRF Internal decoder pipeline circuit for swap control.

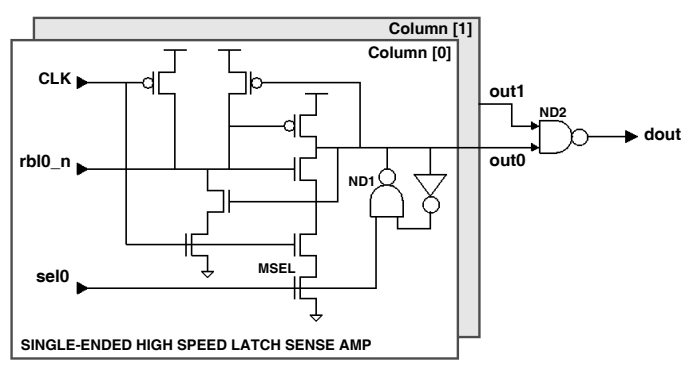


Figure 5.1.5: IRF sense amplifier with built-in 2:1 column decode mux.

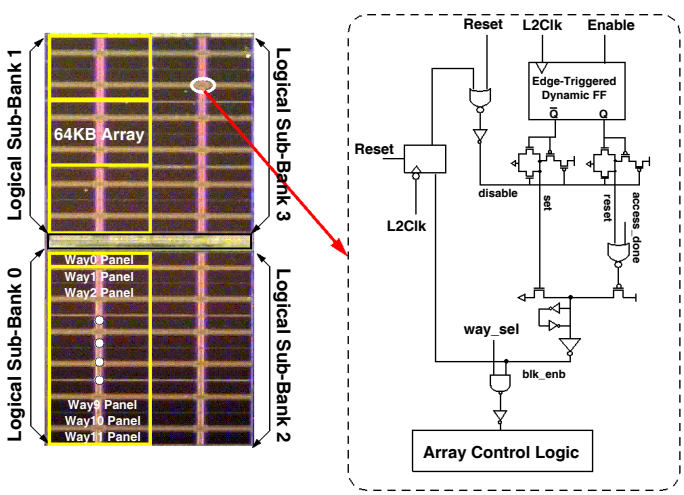
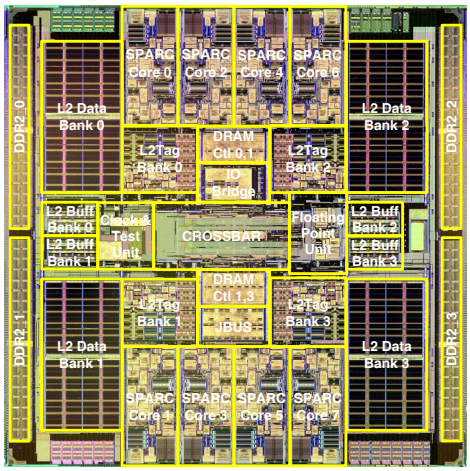


Figure 5.1.6: L2 cache data array floorplan and inter-locking clock header.

Continued on Page 641



Features:

- Eight 64b Multithreaded SPARC Cores
- Shared 3MB L2 Cache
- 16KB ICache per Core
- 8KB DCache per Core
- Four 144b DDR-2 DRAM Interfaces (400 MTs)
- 3.2GB/s JBUS I/O
- Crypto: Public Key (RSA)
- Extensive RAS

Technology:

- 90nm CMOS Process
- 9LM Copper Interconnect
- Power: 63 Watts @ 1.2GHz
- Die Size: 378mm^2
- 279M Transistors
- Package: Flip-chip ceramic LGA (1933 pins)

Figure 5.1.7: Niagara processor micrograph and overview.

